

YoungDeborah540Milestone1

April 9, 2023

0.1 Milestone 1

Project Subject Area: Describe your project in 1-2 sentences

This project is focused around opioid use in the US. I plan to compile information regarding opioid-related overdoses, opioid prescription dispense rates, and demographic data by state.

Flat File: Description: “This data presents provisional counts for drug overdose deaths based on a current flow of mortality data in the National Vital Statistics System.” Link: <https://data.cdc.gov/NCHS/VSRR-Provisional-Drug-Overdose-Death-Counts/xkb8-kh2a>

API: Description: “The American Community Survey (ACS) is an ongoing survey that provides data every year – giving communities the current information they need to plan investments and services. The ACS covers a broad range of topics about social, economic, demographic, and housing characteristics of the U.S. population. Much of the ACS data provided on the Census Bureau’s Web site are available separately by age group, race, Hispanic origin, and sex. Detailed Tables, Subject Tables, Data Profiles, Comparison Profiles and Selected Population Profiles are available for the nation, all 50 states, the District of Columbia, Puerto Rico, every congressional district, every metropolitan area, and all counties and places with populations of 65,000 or more.” Link: <https://www.census.gov/data/developers/data-sets/acs-1year.html> My key: 019d49be4f122d25e6a8216ab35411f7b3b85efa

Website: Description: “The data in the maps show the geographic distribution in the United States, at both state and county levels, of retail opioid prescriptions dispensed per 100 persons per year from 2006–2020” Link: <https://www.cdc.gov/drugoverdose/rxrate-maps/county2020.html>

Relationships - Describe how the data from each source is connected.

Each source is connected by location (state) and timeframe (year).

250 Words describing how you plan to tackle the project, what the data means, the ethical implications of your project scenario/topic, and what challenges you might face.

I recognize that this may be a hefty lift for a class project, but the information regarding opioid prescription, opioid-related mortality, and socio-economic factors is of great interest to me. As I clean and prepare the data, I’ll have to make decisions on how to ethically handle the data. Luckily, the sources I have chosen have already taken into consideration the concern for privacy and confidentiality of the subjects, so there are no identifiers available to me. The data is all publicly available through government websites (CDC and Census Bureau), so I feel confident that there are no ownership issues. The biggest challenge I will face is in what to include and exclude in compiling the sets together. Given that there are likely going to be cells without information, I

will have to decide whether to remove the whole row, or replace those cells with a different value. There will be many columns that I may not know the significance of, so I will have to be sure they are not potentially important for analysis later before removing any. The data is all very different, so I will have to have a clear focus on what output I desire before making any data transformations. Overall, I hope that I can put together a (novice) dataset with a framework that could be applied in a public health intervention. Ideally, I could use it to draw correlations between socioeconomic status of a population, prevalence of opioids via prescription in a region, and per capita mortality rate from opioid-related overdoses. In this ideal scenario, this would provide insight into locations for greater opioid care (care clinics, recovery options, community education).